



Pengembangan Instrumen Tes *Higher Order Thinking Skills (HOTS)* Konten Perbandingan untuk Siswa Fase D

Ulfa Nabilah ^{1*}, Armis ², Sehatta Saragih ³

^{1,2,3} Universitas Riau, Indonesia

*Corresponding Author: ulfa.nabilah0877@student.unri.ac.id

Submitted: 26 November 2025 | Revised: 19 December 2025 | Accepted: 20 December 2025

Abstrak

Kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills (HOTS)* adalah keterampilan penting yang harus dimiliki oleh siswa. Banyak siswa yang memiliki HOTS yang rendah. Penelitian ini bertujuan untuk mengembangkan instrumen tes *Higher Order Thinking Skills (HOTS)* pada konten perbandingan untuk siswa fase D. Penelitian menggunakan model *development studies* dengan dua tahap utama, yaitu *preliminary* dan *formative evaluation*. Tahap *formative evaluation* terdiri dari beberapa langkah yang meliputi *self evaluation*, *expert review*, *one-to-one*, *small group*, dan *field test*. Subjek penelitian melibatkan 30 siswa fase D. Data dikumpulkan melalui lembar validasi, angket keterbacaan siswa, serta hasil tes HOTS, kemudian dianalisis dengan bantuan *software* Anates Uraian 4.0.5. Hasil penelitian menunjukkan bahwa 10 butir soal HOTS yang dikembangkan memenuhi kriteria valid secara rasional maupun empiris, memiliki reliabilitas sangat tinggi (0,95), tingkat kesukaran yang bervariasi (dua mudah, enam sedang, dua sukar), serta daya pembeda yang baik. Dengan demikian, Instrumen tes HOTS konten perbandingan ini layak digunakan sebagai alat evaluasi untuk melatih keterampilan berpikir kritis, analitis, dan kreatif siswa.

Kata Kunci: HOTS, pengembangan instrumen, perbandingan, asesmen, siswa fase D

Abstract

Higher Order Thinking Skills (HOTS) are essential competencies that students must possess. Many students still exhibit low levels of HOTS. This study aims to develop a Higher Order Thinking Skills (HOTS) instrument in comparison content for Phase D students. The research employed a development studies model consisting of two main stages: the preliminary stage and formative evaluation. The formative evaluation stage comprised several steps, including self-evaluation, expert review, one-to-one, small group, and field testing. The research subjects involved 30 Phase D students. Data were collected through validation sheets, student readability questionnaires, and HOTS test results, which were analyzed using Anates Uraian 4.0.5 software. The results indicated that the ten HOTS items developed met both rational and empirical validity criteria, showed very high reliability (0.95), had varying levels of difficulty (two easy, six moderate, and two difficult items), and demonstrated good discriminating power. Therefore, the developed HOTS instrument on comparison content is considered feasible to be used as an evaluation tool to enhance students' critical, analytical, and creative thinking skills.

Keywords: HOTS, instrument development, comparison, assessment, Phase D students

PENDAHULUAN

Keterampilan berpikir tingkat tinggi atau *Higher Order Thinking Skills (HOTS)* memiliki peran sentral dalam pendidikan modern karena mencakup kemampuan analisis, evaluasi, dan kreativitas untuk memecahkan persoalan yang kompleks (Tim Pusat Penelitian, 2019). Siswa dituntut tidak hanya menghafal informasi, melainkan memahami konsep secara mendalam serta mampu bernalar kritis dan logis ketika menyelesaikan permasalahan (Suyatno, dkk., 2023). Tujuan penerapan HOTS adalah mengembangkan kemampuan berpikir kritis, kreatif, dan inovatif yang selaras dengan tuntutan kehidupan nyata (Direktorat Jenderal Guru dan Tenaga Kependidikan, 2019).



Pembelajaran matematika menuntut keterlibatan HOTS secara optimal agar siswa mampu menghadapi tantangan abad 21. Penerapan soal-soal HOTS terbukti efektif mendorong siswa berpikir di level kognitif C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta) sebagaimana diuraikan dalam Taksonomi Bloom Revisi (Anderson & Krathwohl dalam Tim Pusat Penelitian, 2019). Penelitian sebelumnya menyebutkan bahwa soal berbasis HOTS mampu meningkatkan keterampilan berpikir kritis sekaligus efektivitas pembelajaran (Gradini, 2019). Temuan empiris menunjukkan bahwa keterampilan siswa dalam menyelesaikan soal HOTS masih tergolong rendah. Studi yang dilakukan Tania (2021) melaporkan persentase rata-rata kemampuan siswa pada level C4 hanya mencapai 53,12% (kategori rendah), C5 sebesar 67,2% (kategori sedang), dan C6 sekitar 54,98% (kategori rendah). Fakta ini mengindikasikan perlunya intervensi untuk meningkatkan keterampilan berpikir tingkat tinggi melalui penyediaan instrumen soal yang memadai.

Asesmen membantu siswa terbiasa menghadapi soal-soal yang menuntut pemikiran tinggi dan pada saat yang sama membantu guru mengetahui tingkat penguasaan kognitif siswa (Wulandari, dkk., 2020). Kurikulum merdeka menyediakan tiga jenis asesmen yang dapat digunakan guru, yaitu diagnostik untuk memetakan kemampuan awal siswa, formatif untuk memantau perkembangan pembelajaran, serta sumatif untuk mengevaluasi capaian akhir (Kemendikbud, 2025).

Hasil penelitian Kurniasi & Arsisari (2020) mengungkapkan bahwa sebagian besar guru masih mengandalkan soal dari buku paket yang belum sepenuhnya melatih keterampilan berpikir tingkat tinggi. Pusparini & Mistiani (2023) merekomendasikan penyusunan soal berbasis konteks nyata agar siswa terbiasa berpikir kritis, analitis, dan kreatif. Minimnya ketersediaan soal HOTS mengakibatkan siswa kurang mendapatkan kesempatan untuk mengasah kemampuan berpikir tingkat tinggi secara optimal. Konten perbandingan dalam matematika memiliki keterkaitan erat dengan situasi kehidupan sehari-hari, misalnya membandingkan harga barang, waktu perjalanan, atau skala peta (Wahyuni & Prihatiningtyas, 2020). Penelitian Ummah & Khotimah (2021) menyatakan bahwa latihan soal kontekstual pada topik ini dapat mendorong siswa berpikir lebih logis dan mendalam. Wawancara dengan guru di beberapa sekolah mengindikasikan keterbatasan soal HOTS pada materi perbandingan, sehingga siswa belum terbiasa menghadapi persoalan kontekstual yang membutuhkan pemikiran kritis dan kreatif.

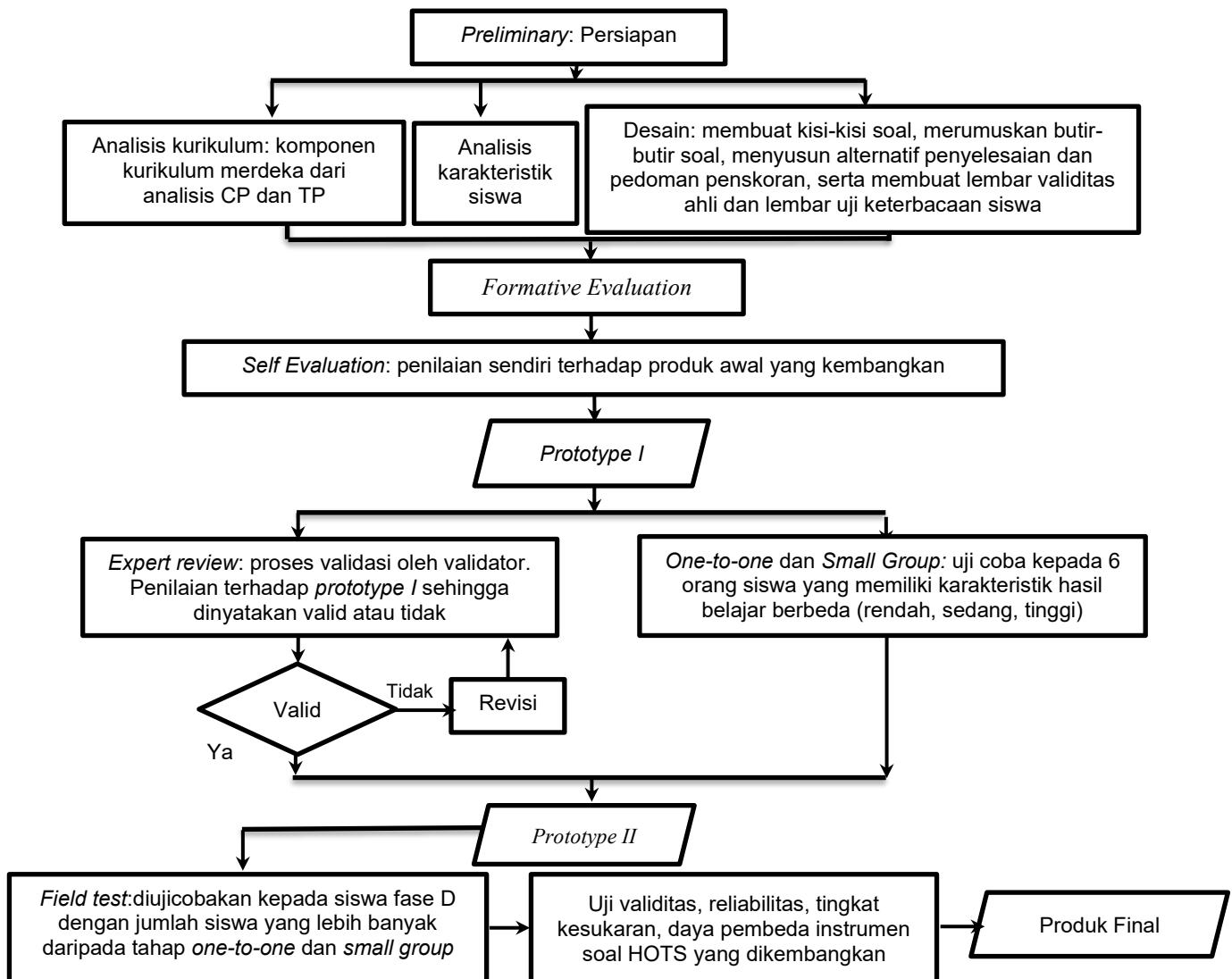
Penelitian Panjaitan, dkk. (2022) melaporkan bahwa 33% siswa mengalami kesulitan menyelesaikan soal cerita pada konten perbandingan. Hoar, dkk. (2021) juga menegaskan bahwa rendahnya keseriusan siswa dalam mempelajari konten perbandingan serta kesalahan mendasar yang muncul memengaruhi capaian pembelajaran. Bukti tersebut memperkuat pentingnya pengembangan instrumen soal yang dapat memfasilitasi keterampilan berpikir tingkat tinggi secara komprehensif.

Instrumen yang berkualitas perlu memenuhi kriteria validitas, reliabilitas, tingkat kesukaran yang proporsional, serta daya pembeda yang memadai (Wati, 2018). Konsep validitas berkaitan dengan ketepatan instrumen dalam mengukur kompetensi yang diharapkan, sedangkan reliabilitas menekankan pada konsistensi hasil pengukuran. Tingkat kesukaran dan daya pembeda menentukan kemampuan soal dalam mengklasifikasikan siswa berdasarkan tingkat pemahaman yang berbeda (Yusuf, 2015).

Pengembangan instrumen soal HOTS untuk topik perbandingan pada siswa fase D diharapkan mampu menjawab tantangan tersebut. Penyediaan soal yang memenuhi kriteria kualitas asesmen akan membantu guru dalam melaksanakan evaluasi pembelajaran yang sesuai dengan tuntutan kurikulum merdeka sekaligus melatih keterampilan berpikir tingkat tinggi siswa. Ketersediaan Instrumen tes HOTS yang valid, reliabel, memiliki tingkat kesukaran seimbang, serta daya pembeda yang baik dapat memperbaiki kualitas pembelajaran matematika. Siswa juga akan lebih siap menghadapi soal-soal yang menantang kemampuan berpikir kritis, analitis, dan kreatif sehingga kompetensi abad 21 dapat tercapai secara optimal.

METODE

Penelitian ini menggunakan model pengembangan *development studies* yang terdiri dari dua tahapan pokok. Tahap pertama adalah *preliminary*, yang meliputi kegiatan analisis kebutuhan dan perancangan instrumen. Tahap kedua yaitu *formative evaluation*, yang disusun berdasarkan model Tessmer dan mencakup beberapa langkah, antara lain *self evaluation*, *expert review*, *one-to-one*, *small group*, serta *field test* (Heriyadi & Prahmana, 2020). Prosedur pengembangan secara keseluruhan dapat digambarkan sebagai berikut.



Gambar 1. Bagan prosedur pengembangan

Tahap *preliminary* mencakup kegiatan analisis terhadap kebutuhan, kurikulum dan siswa, serta desain produk awal dan instrumen penelitian. Pada tahap *formative evaluation*, ada beberapa langkah yang dilaksanakan. Pertama, *self evaluation*, yaitu menilai dan memperbaiki produk awal. Hasil dari *self evaluation* disebut *prototype I*. *Prototype I* kemudian divalidasi oleh validator dalam tahap *expert review*, menilai dari aspek materi, konstruk, dan bahasa. Jika produk dinyatakan valid, tahap berikutnya dapat dilanjutkan. Jika belum valid, revisi dilakukan berdasarkan saran validator.

Tahap selanjutnya adalah tahap *one-to-one* dan *small group* yang dilakukan setelah revisi hasil *expert review*. Pada tahap ini, *prototype I* diujicobakan kepada enam siswa fase D dengan kemampuan heterogen yang telah mempelajari konten perbandingan. Siswa diminta mengerjakan dan memberikan komentar terkait kejelasan, kelengkapan, dan keterbacaan soal. Siswa juga diminta mengisi angket uji keterbacaan siswa. Komentar siswa pada tahap ini menjadi acuan untuk memperbaiki *prototype I*. Masukan dari siswa digunakan untuk menyempurnakan instrumen sehingga menghasilkan *prototype II*.

Tahap terakhir yaitu *field test* melibatkan subjek penelitian yang diujicobakan kepada 30 siswa fase D. Data diperoleh melalui hasil tes dan angket respon siswa, kemudian dianalisis menggunakan *software* Anates Uraian 4.0.5 untuk menentukan validitas butir soal, reliabilitas, tingkat kesukaran, dan daya pembeda. Produk Nakhir yang memenuhi kriteria kualitas instrumen kemudian ditetapkan sebagai soal HOTS final yang siap digunakan dalam pembelajaran.

Data dalam penelitian ini diperoleh secara rasional melalui lembar validasi dan angket uji keterbacaan siswa, serta secara empiris melalui soal. Data yang menggunakan skala Likert (seperti lembar validasi dan angket uji keterbacaan siswa) terlebih dahulu diubah dari data kualitatif berupa sangat setuju/sesuai (SS), setuju/sesuai (S), kurang setuju/sesuai (KS), dan tidak setuju/sesuai (TS) menjadi data kuantitatif dengan ketentuan skor: 4 untuk SS, 3 untuk S, 2 untuk KS, dan 1 untuk TS.

Pengumpulan data dilakukan menggunakan teknik angket dan tes, serta instrumen penelitian berupa lembar validasi, angket respon siswa dan soal HOTS. Lembar validasi diisi oleh tiga validator pada tahap *expert review* untuk memperoleh hasil validitas rasional dari aspek materi, konstruk dan bahasa, serta untuk mendapatkan komentar dan saran perbaikan terhadap instrumen soal yang dikembangkan. Angket uji keterbacaan siswa dikerjakan oleh siswa pada tahap *one-to-one* dan *small group*, bertujuan melihat respon siswa terhadap keterbacaan instrumen soal HOTS. Soal HOTS diberikan untuk mendapatkan jawaban siswa pada tahap *field test*, yang diperlukan dalam menganalisis validitas empiris, reliabilitas, tingkat kesulitan, dan daya pembeda.

Validitas rasional adalah uji kevalidan kepada para ahli/validator, sedangkan validitas empiris adalah kevalidan yang diuji kepada siswa dengan melihat reliabilitas, tingkat kesulitan soal, dan daya pembedanya (Riyani, dkk., 2017). Untuk mengetahui kevalidan instrumen soal HOTS dilakukan perhitungan nilai validasi dengan menggunakan bantuan *Microsoft Excel*. Penilaian dari validator berupa skala likert (sangat sesuai, sesuai, kurang sesuai, tidak sesuai) diubah ke dalam bentuk skor 4, 3, 2 dan 1. Untuk menginterpretasi nilai validitas rasional yang diperoleh, selanjutnya digunakan pengklasifikasian validitas menurut pendapat Ratumanan & Laurens (dalam Mita, dkk., 2023) dengan ketentuan seperti yang disajikan pada tabel 1 berikut.

Tabel 1. Kriteria Validitas Rasional

Nilai rata-rata total validitas	Tingkat Validitas
$3,5 \leq RTV \leq 4,0$	Sangat Valid
$2,5 \leq RTV < 3,5$	Valid
$1,5 \leq RTV < 2,5$	Kurang Valid
$1,0 \leq RTV < 1,5$	Tidak Valid

(Modifikasi dari Ratumanan & Laurens dalam Mita, dkk., 2023)

Instrumen soal HOTS dikatakan valid jika RTV lebih besar dari 2,5.

Angket uji keterbacaan siswa diberikan kepada siswa setelah selesai mengerjakan instrumen soal HOTS. Untuk mengetahui nilai keterbacaan atau respon siswa terhadap instrumen soal HOTS dilakukan perhitungan dengan menggunakan bantuan *Microsoft Excel*.. Hasil penilaian dari siswa berupa skala likert (sangat setuju, setuju, kurang setuju, tidak setuju) diubah ke dalam bentuk skor 4, 3, 2 dan 1. Angket uji keterbacaan siswa dianalisis dan diinterpretasikan menurut pendapat Vebrian, dkk. (2022) dengan ketentuan seperti pada tabel 2 sebagai berikut.

Tabel 2. Kriteria Uji keterbacaan

Percentase	Kriteria
$81,25\% \leq P \leq 100\%$	Sangat baik
$62,5\% \leq P < 81,25\%$	Baik
$43,75\% \leq P < 62,5\%$	Kurang baik
$25\% \leq P < 43,75\%$	Tidak baik

(Modifikasi dari Vebrian, dkk., 2022)

Respon siswa terhadap instrumen asesmen dikatakan baik apabila persentase lebih dari 50%.

Pada tahap uji validitas empiris, data diperoleh melalui pelaksanaan uji coba instrumen tes kepada subjek penelitian, kemudian dilakukan analisis terhadap validitas, reliabilitas, tingkat kesukaran, serta daya pembeda butir soal. Pengujian validitas dilakukan menggunakan rumus korelasi *Product Moment* dengan bantuan *software Anates Uraian* versi 4.0.5. Nilai r_{hitung} yang diperoleh kemudian dibandingkan dengan nilai r_{tabel} pada taraf signifikansi $\alpha = 5\%$ dan jumlah responden (n) sebanyak 30 siswa. Suatu butir soal dinyatakan valid apabila memenuhi kriteria $r_{hitung} \geq r_{tabel}$.

Uji reliabilitas dilakukan dengan menerapkan rumus *Alpha Cronbach* menggunakan *software Anates Uraian* versi 4.0.5. Suatu instrumen dapat dinyatakan reliabel apabila nilai koefisien reliabilitasnya $\geq 0,70$ (Mansyur, dkk., 2019).

Analisis tingkat kesukaran dilakukan dengan menggunakan rumus tingkat kesukaran yang diolah dengan bantuan *software Anates Uraian* versi 4.0.5. Suatu butir soal dianggap memiliki tingkat kesukaran yang baik apabila tidak terlalu mudah maupun terlalu sukar (Lestari & Yudhanegara, 2018). Soal dikatakan sangat mudah apabila hampir seluruh siswa menjawab benar, sedangkan sangat sukar apabila sebagian besar siswa menjawab salah.

Uji daya pembeda dilakukan dengan mengurutkan skor siswa dari tertinggi hingga terendah, kemudian membaginya menjadi dua kelompok, yaitu kelompok atas 27% dan kelompok bawah 27%. Setelah data kedua kelompok diperoleh, dilakukan perhitungan daya pembeda menggunakan rumus yang sesuai dengan bantuan *software Anates Uraian* versi 4.0.5. Butir soal yang dapat diterima sebagai instrumen adalah yang memiliki daya pembeda dengan kategori sangat baik, baik, atau cukup (Lestari & Yudhanegara, 2018).

HASIL PENELITIAN

Produk akhir yang dihasilkan pada penelitian pengembangan ini berupa instrumen Tes *Higher Order Thinking Skills* (HOTS) konten perbandingan untuk siswa fase D yang memiliki kriteria valid secara rasional, valid secara empiris, reliabilitas yang tinggi, tingkat kesukaran dan daya pembeda yang baik. Pengembangan instrumen soal ini dilakukan dengan mengikuti prosedur pengembangan tipe *development study* sebagai berikut:

Tahap *Preliminary*

Tahapan *preliminary* mencakup kegiatan analisis yang dilakukan peneliti sebagai langkah awal dalam proses pengembangan instrumen soal HOTS. Analisis awal yang dilaksanakan peneliti berupa analisis kebutuhan. Berdasarkan hasil analisis kebutuhan diperoleh informasi dari hasil wawancara yang dilakukan kepada dua guru matematika di dua sekolah berbeda yang menyatakan bahwa guru cukup kesulitan untuk memberikan soal HOTS dalam melatih siswa. Hal ini dikarenakan guru kurang referensi untuk membuat soal HOTS, sehingga guru hanya memberikan soal-soal yang ada dibuku paket matematika di sekolah saja. Jenis soal-soal tes pada buku paket yang digunakan guru juga belum teruji kevalidan, reliabilitas, tingkat kesukaran dan daya pembeda dalam melatih HOTS siswa.

Hasil analisis kurikulum menunjukkan tujuan pembelajaran yang selaras dengan capaian pembelajaran (CP) pada konten perbandingan dalam kurikulum merdeka. Capaian pembelajaran konten perbandingan di fase D pada elemen bilangan dengan capaian pembelajaran sebagai berikut: "Siswa dapat menggunakan rasio (skala, proporsi, dan laju perubahan) dalam penyelesaian masalah". Berdasarkan capaian pembelajaran tersebut peneliti membuat tujuan pembelajaran terfokus pada konten perbandingan yaitu: 1) menyelesaikan permasalahan yang berkaitan dengan rasio dua besaran yang satuanya sama; 2) menyelesaikan permasalahan yang berkaitan dengan perbandingan senilai dengan menggunakan tabel data; 3) menyelesaikan permasalahan kontekstual yang berkaitan dengan perbandingan senilai; 4) menyelesaikan permasalahan kontekstual yang berkaitan dengan perbandingan berbalik nilai; 5) menyelesaikan permasalahan yang berkaitan dengan perbandingan berbalik nilai dengan menggunakan tabel data; 6) mengevaluasi skala berbeda secara numerik dan logis serta menyimpulkan pilihan yang paling tepat dalam konteks dunia nyata. Pada analisis siswa diperoleh bahwa ujicoba soal tes dilakukan kepada siswa fase D yang telah mempelajari konten perbandingan.

Berdasarkan hasil analisis yang telah dilakukan, peneliti membuat rancangan produk awal yang terdiri dari kisi-kisi soal, butir soal, alternatif penyelesaian, pedoman penskoran. Peneliti juga merancang instrumen penelitian yang terdiri atas lembar validasi dan angket uji keterbacaan siswa.

Tahap *Formative Evaluation*.

Untuk langkah-langkah pada tahap *formative evaluation* diadopsi dari Tessmer terdiri atas tahapan sebagai berikut:

Self Evaluation

Tahap *self evaluation* dilaksanakan dengan meninjau dan memeriksa secara cermat kelengkapan produk awal yang telah dirancang oleh peneliti. Perbaikan pada tahap ini meliputi penyesuaian struktur kalimat serta penyempurnaan dalam penggunaan simbol-simbol matematika. Berdasarkan hasil penelaahan pada tahap *self evaluation*, diperoleh 10 butir soal

yang dinyatakan siap untuk divalidasi oleh validator. Produk yang dihasilkan pada tahap ini disebut sebagai *prototype I*.

Expert Review

Prototype I kemudian divalidasi oleh tiga orang validator untuk memperoleh data mengenai validitas rasional yang mencakup tiga aspek, yaitu materi, konstruk, dan bahasa. Proses penilaian dilakukan oleh para validator dengan mengisi lembar validasi menggunakan *skala Likert* empat tingkat, yakni sangat sesuai, sesuai, kurang sesuai, dan tidak sesuai. Setiap kategori penilaian tersebut kemudian dikonversi ke dalam bentuk data kuantitatif dengan ketentuan skor: 4 untuk sangat sesuai, 3 untuk sesuai, 2 untuk kurang sesuai, dan 1 untuk tidak sesuai. Berdasarkan hasil penilaian dari ketiga validator, diperoleh rata-rata skor validitas rasional instrumen soal HOTS yang dapat dilihat pada tabel 3 berikut.

Tabel 3. Rekapitulasi Hasil Validitas Rasional Soal HOTS

Aspek Penilaian	Rata-Rata Skor Penilaian Pervalidator			Rata-Rata Tiap Aspek
	V1	V2	V3	
Materi	4	3.7	3.9	3.9
Konstruk	4	3.3	3.7	3.7
Bahasa	3.8	3.7	3.9	3.8
Rata-Rata Total Validitas (RTV)				3.8
Kriteria				Sangat Valid

Rata-rata validitas yang diberikan validator, aspek konstruk untuk soal HOTS memperoleh nilai terendah karena tedapat beberapa kesalahan kalimat yang belum sesuai dengan rumusan pokok item dalam penulisan soal. Peneliti tidak memberikan gambar ilustrasi untuk soal nomor 2, 5 dan 8 sehingga disarankan untuk menambahkan gambar ilustrasi. Rata-rata keseluruhan nilai validitas dari ketiga aspek tersebut sebesar 3,8 dan termasuk dalam kategori sangat valid, sehingga dapat disimpulkan bahwa instrumen soal HOTS yang dikembangkan telah memenuhi kriteria validitas rasional atau teoretis.

One-to-One dan Small Group

Prototype I yang telah dikembangkan diuji coba terhadap enam siswa dengan tingkat kemampuan matematis bervariasi, meliputi tinggi, sedang, dan rendah, melalui tahap *one-to-one* sekaligus *small group*. Tahap ini menugaskan siswa untuk menyelesaikan soal-soal HOTS yang telah disusun sebelumnya. Hasil analisis respons siswa menunjukkan bahwa indikator HOTS yang diharapkan mulai tampak dalam proses tersebut. Jawaban siswa memperlihatkan kemampuan memahami serta menuliskan pernyataan matematis secara tepat tanpa mencantumkan unsur diketahui maupun ditanyakan dalam soal. Langkah berikut menunjukkan bahwa siswa belum mampu menguraikan jawaban secara runtut dan sistematis. Pertanyaan tersebut sudah dapat disimpulkan dengan keputusan yang sesuai konteks permasalahan.

Selesai mengerjakan instrumen soal HOTS, peneliti meminta siswa mengisi angket uji keterbacaan. Tujuan kegiatan tersebut untuk mengetahui respon siswa terhadap instrumen yang dikembangkan. Rata-rata persentase respon siswa terhadap soal HOTS mencapai 82,73%, menunjukkan kategori sangat baik terhadap instrumen yang telah dikembangkan peneliti. Saran dan komentar siswa menjadi dasar revisi *prototype II* yang mencakup

penambahan kejelasan informasi soal serta penyederhanaan rumusan kalimat pada butir pertanyaan.

Field Test

Tahap *field test* melibatkan pengujian *prototype* // terhadap subjek penelitian. Subjek uji coba terdiri atas 30 siswa dalam fase D. Jawaban siswa dalam tahap *field test* diberi skor kemudian dianalisis untuk memperoleh data mengenai validitas empiris, reliabilitas, tingkat kesukaran, serta daya pembeda butir soal. Proses uji validitas butir soal menggunakan *software* Anates Uraian versi 4.0.5 terhadap 10 butir soal. Setelah dilakukan pengujian, berikut hasil uji validitas empiris yang dapat dilihat pada tabel 4 berikut.

Tabel 4. Hasil Uji Validitas Empiris

No Butir	Level Kognitif	r_{hitung}	r_{tabel}	Keputusan
1	C5	0,801	0,361	Valid
2	C5	0,671		Valid
3	C4	0,841		Valid
4	C6	0,733		Valid
5	C4	0,781		Valid
6	C5	0,623		Valid
7	C4	0,864		Valid
8	C5	0,778		Valid
9	C5	0,743		Valid
10	C6	0,865		Valid

Berdasarkan hasil analisis validitas butir soal yang dilakukan diperoleh sebanyak 10 butir soal yang termasuk dalam kategori valid.

Nilai reliabilitas untuk soal HOTS mencapai 0,95, yang menunjukkan kriteria sangat tinggi. Mengacu pada pendapat Mansyur, dkk. (2019), suatu instrumen dapat dinyatakan reliabel apabila memiliki nilai reliabilitas minimal 0,70, sehingga instrumen soal HOTS yang dikembangkan telah memenuhi syarat reliabilitas. Dengan demikian, soal HOTS yang disusun memiliki tingkat keakuratan dan konsistensi jawaban yang tinggi, meskipun diuji pada subjek, lokasi, atau waktu yang berbeda.

Hasil dari analisis dan perhitungan uji tingkat kesukaran pada penelitian ini untuk 10 soal bentuk uraian dihitung secara otomatis menggunakan *software* Anates Uraian versi 4.0.5 dapat dilihat pada tabel 5 berikut.

Tabel 5. Hasil Uji Tingkat Kesukaran

No Butir	Level Kognitif	Tkt. Kesukaran (%)	Tafsiran
1	C5	68,23	Sedang
2	C5	67,71	Sedang
3	C4	54,30	Sedang
4	C6	28,52	Sukar
5	C4	71,88	Mudah
6	C5	37,50	Sedang
7	C4	29,69	Sukar
8	C5	71,35	Mudah
9	C5	34,90	Sedang
10	C6	39,45	Sedang

Analisis terhadap tingkat kesukaran soal HOTS dilakukan untuk memastikan adanya keseimbangan proporsi tingkat kesulitan pada butir soal. Dari 10 soal yang dikembangkan, diperoleh dua soal kategori mudah, enam soal kategori sedang, dan dua soal kategori sukar. Seluruh soal tersebut dinyatakan layak digunakan karena tidak terdapat butir yang tergolong terlalu mudah maupun terlalu sukar. Temuan ini sejalan dengan pendapat Lestari & Yudhanegara (2018) yang menyatakan bahwa soal dengan tingkat kesukaran yang baik adalah soal yang tidak terlalu sulit atau terlalu mudah.

Hasil analisis dan perhitungan daya pembeda soal HOTS dihitung secara otomatis menggunakan *software* Anates Uraian versi 4.0.5 dapat dilihat pada tabel 6 berikut.

Tabel 6. Hasil Uji Daya Pembeda

No Butir Soal	Level Kognitif	DP (%)	Kriteria
1	C5	40,63	Baik
2	C5	50,00	Baik
3	C4	67,97	Baik
4	C6	30,47	Cukup
5	C4	32,81	Cukup
6	C5	60,42	Baik
7	C4	50,00	Baik
8	C5	44,79	Baik
9	C5	65,63	Baik
10	C6	61,72	Baik

Hasil analisis terhadap daya pembeda soal HOTS menunjukkan bahwa terdapat delapan soal memiliki daya pembeda baik dan dua soal tergolong cukup. Seluruh butir soal yang dikembangkan layak digunakan, karena memiliki nilai daya pembeda di atas 20%, yang menandakan bahwa setiap butir mampu membedakan kemampuan siswa secara efektif. Dengan demikian, instrumen yang dikembangkan dapat menggambarkan perbedaan kemampuan antara siswa berkemampuan tinggi dan rendah.

Tabel 7. Rekapitulasi Analisis Validitas Empiris, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal HOTS

No	Validitas Empiris		Reliabilitas		Tingkat Kesukaran		Daya Pembeda		Keputusan
	r_{hitung}	Kriteria	r_{11}	Kriteria	TK (%)	Kriteria	DP (%)	Kriteria	
1	0,801	Valid	0,95	Sangat Tinggi	68,23	Sedang	40,63	Baik	Diterima
2	0,671	Valid			67,71	Sedang	50,00	Baik	Diterima
3	0,841	Valid			54,30	Sedang	67,97	Baik	Diterima
4	0,733	Valid			28,52	Sukar	30,47	Cukup	Diterima
5	0,781	Valid			71,88	Mudah	32,81	Cukup	Diterima
6	0,623	Valid			37,50	Sedang	60,42	Baik	Diterima
7	0,864	Valid			29,69	Sukar	50,00	Baik	Diterima
8	0,778	Valid			71,35	Mudah	44,79	Baik	Diterima
9	0,743	Valid			34,90	Sedang	65,63	Baik	Diterima
10	0,865	Valid			39,45	Sedang	61,72	Baik	Diterima

PEMBAHASAN

Penelitian ini menghasilkan sebuah instrumen tes HOTS pada konten perbandingan bagi siswa fase D. Pengembangan instrumen dilaksanakan melalui model *development studies* yang mencakup dua tahap utama, yaitu *preliminary* dan *formative evaluation* yang diadaptasi dari rancangan Tessmer. Setiap tahap memiliki peran penting dalam memastikan mutu produk yang dihasilkan, baik dari segi materi, konstruk dan bahasa.

Hasil penelitian menunjukkan bahwa sepuluh butir soal uraian yang dikembangkan telah memenuhi standar kevalidan, baik secara rasional maupun empiris. Berdasarkan penilaian tiga validator ahli, rata-rata skor validitas mencapai 3,8 dengan kategori sangat valid. Aspek materi memperoleh nilai tertinggi, sementara aspek konstruk memerlukan revisi kecil pada kejelasan redaksi dan penggunaan ilustrasi penunjang. Perbaikan dilakukan melalui penyusunan ulang kalimat dan penambahan gambar kontekstual agar makna soal menjadi lebih jelas bagi siswa. Hasil revisi tersebut menjadikan instrumen layak untuk dilanjutkan ke tahap pengujian selanjutnya.

Pada tahap *field test*, diperoleh hasil uji validitas empiris menggunakan *software* Anates Uraian versi 4.0.5 juga memperlihatkan bahwa seluruh butir soal tergolong valid, sehingga setiap soal mampu mengukur aspek kognitif sesuai dengan indikator HOTS yang ditetapkan. Diperoleh juga data bahwa instrumen memiliki tingkat reliabilitas sangat tinggi, yaitu sebesar 0,95, melampaui batas minimal 0,70 sebagaimana dikemukakan oleh Mansyur, dkk. (2019). Nilai tersebut menunjukkan bahwa instrumen yang dikembangkan bersifat konsisten dan dapat digunakan untuk melatih kemampuan berpikir tingkat tinggi secara stabil di berbagai kondisi.

Hasil tingkat kesukaran pada tahap *field test* menunjukkan pola distribusi yang lebih seimbang dibandingkan tahap one-to-one dan small group. Terdapat dua soal kategori mudah (nomor 5 dan 8), enam soal kategori sedang (nomor 1, 2, 3, 6, 9, dan 10), dan dua soal kategori sukar (nomor 4 dan 7). Distribusi ini sudah mendekati proporsi kesukaran yang direkomendasikan Widoyoko (2014), yaitu tidak terlalu mudah dan tidak terlalu sukar. Analisis tingkat kesukaran menunjukkan variasi tingkat kesukaran yang proporsional, yaitu terdapat soal dengan kategori mudah, sedang, dan sukar. Variasi tingkat kesukaran ini penting untuk memberikan tantangan dan kesempatan bagi siswa untuk berpikir tingkat tinggi. Temuan ini mengindikasikan bahwa instrumen telah memberikan peluang bagi seluruh siswa untuk berusaha mencapai penyelesaian optimal.

Daya pembeda soal menunjukkan bahwa sebagian besar butir soal berada pada kategori baik, yaitu delapan soal berkategori baik dan dua soal berkategori cukup. Daya pembeda yang baik menunjukkan bahwa instrumen mampu membedakan siswa dengan kemampuan tinggi dan rendah. Temuan ini relevan dengan pendapat Lestari & Yudhanegara (2018) yang menyebutkan bahwa soal dengan daya pembeda lebih besar atau sama dengan 20% dapat diterima karena mampu mengidentifikasi kemampuan siswa secara akurat.

Secara keseluruhan, hasil pengembangan membuktikan bahwa Instrumen tes HOTS konten perbandingan yang dihasilkan memenuhi kriteria kevalidan, reliabilitas, tingkat kesukaran, serta daya pembeda yang baik, tetapi masih terdapat keterbatasan dalam keseimbangan tingkat kesukaran dan distribusi subtopik antarbutir soal. Penelitian berikutnya diharapkan dapat memperluas cakupan submateri dan menyesuaikan proporsi soal agar lebih representatif terhadap keseluruhan elemen pembelajaran matematika pada fase D, sehingga

instrumen yang dihasilkan dapat digunakan secara lebih luas dalam konteks asesmen pembelajaran.

SIMPULAN DAN SARAN

Penelitian ini menerapkan model pengembangan tipe *development studies* yang mencakup dua tahap utama, yaitu tahap pendahuluan (*preliminary*) dan tahap evaluasi formatif (*formative evaluation*). Hasil akhir dari kegiatan pengembangan ini berupa instrumen tes kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS) yang terdiri atas sepuluh butir soal uraian. Instrumen tes HOTS pada konten perbandingan untuk siswa fase D terbukti telah memenuhi kriteria kevalidan, baik secara teoritis maupun empiris, serta memiliki reliabilitas tinggi, daya pembeda memadai, dan tingkat kesulitan yang baik. Sebagai tindak lanjut, peneliti selanjutnya diharapkan dapat mengembangkan instrumen sejenis dengan memperhatikan komposisi tingkat kesukaran yang lebih proporsional, agar diperoleh alat asesmen yang lebih representatif dan akurat dalam melatih kemampuan siswa.

DAFTAR PUSTAKA

- Direktorat Jenderal Guru dan Tenaga Kependidikan. (2019). *Buku Penilaian Berorientasi Higher Order Thinking Skills*. Kementerian Pendidikan dan Kebudayaan.
- Gradini, E. (2019). Menilik Konsep Kemampuan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) dalam Pembelajaran Matematika. *Jurnal Numeracy*, 6(2).
- Heriyadi, & Prahmana, R. C. I. (2020). Pengembangan Lembar Kegiatan Siswa Menggunakan Pendekatan Pendidikan Matematika Realistik. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(2). <https://doi.org/10.24127/ajpm.v9i2.2782>
- Hoar, A. Y., Amsikan, S., & Nahak, S. (2021). Analisis Kesalahan Siswa dalam Menyelesaikan Soal Cerita Perbandingan. *Jurnal Ilmu Pendidikan Matematika*, 6(1).
- Kemendikbud. (2025). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah Edisi Revisi Tahun 2025. Badan Standar, Kurikulum, dan Asesmen Pendidikan (BSKAP)*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Kurniasi, E. R., & Arsisari, A. (2020). Pengembangan Instrumen Pengukur Higher Order Thinking Skills (HOTS) Matematika Pada Siswa Sekolah Menengah Pertama. *Jurnal Program Studi Pendidikan Matematika*, 9(4). <https://doi.org/10.24127/ajpm.v9i4.3162>
- Lestari, K. E., & Yudhanegara, M. R. (2018). *Penelitian Pendidikan Matematika*.
- Mansyur, Rasyid, H., & Suratno. (2019). *Asesmen Pembelajaran di Sekolah Panduan bagi Guru dan Calon Guru*. Pustaka Belajar.
- Mita., Manfaat, B., & Izzati, N. (2023). Pengembangan Instrumen Penilaian Berbasis Proyek untuk Membentuk Kreativitas Motorik Matematik Siswa SMP pada Pokok Bahasan Transformasi Geometri. *Teorema: Teori Dan Riset Matematika*, 8(2). <https://doi.org/10.25157/teorema.v8i2.12111>
- Panjaitan, & dkk. (2022). Analisis Kesulitan Peserta Didik Menyelesaikan Soal Cerita Pada Materi Perbandingan Kelas VII UPT SMP Negeri 37 Medan. *SEPREN: Journal of Mathematics Education and Applied*, 03(02).

- Pusparini, F., & Mistiani, L. (2023). Profil Kemampuan HOTS Siswa SMP Plus Ma'arif NU Ciamis pada Pelajaran Biologi. *Bioed: Jurnal Pendidikan Biologi*, 11(2). <https://doi.org/10.25157/jpb.v11i2.11926>
- Riyani, R., & dkk. (2017). Uji Validitas Pengembangan Tes untuk Mengukur Kemampuan Pemahaman Rasional pada Materi Persamaan Kuadrat Siswa Kelas VIII SMP. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, 1(1).
- Suyatno, Juharni, I., & Susilowati, W. W. (2023). *Teori Belajar & Pembelajaran Berorientasi Higher Order Thinking Skills*. K-Media.
- Tania, D. (2021). *Analisis Kemampuan HOTS (Higher Order Thinking Skills) Siswa pada Pembelajaran Online di SMAN 1 Teluk Kuantan Tahun Pelajaran 2020/2021* [Skripsi].
- Tim Pusat Penelitian. (2019). *Panduan Penulisan Soal HOTS (Higher Order Thinking Skills)*. Pusat Penilaian Pendidikan.
- Ummah, W. C. C., & Khotimah, R. P. (2021). Keterampilan Berpikir Kritis Siswa dalam Menyelesaikan Soal Aritmatika Sosial Berbasis HOTS di Kelas VII MTs. *Khozinatul 'Ulum. Jurnal Varidika*, 33(2).
- Vebrarian, R., Putra, Y. Y., & Saraswati, S. (2022). Respon Siswa dalam Menyelesaikan Soal Literasi Matematika pada Konteks Bangka Belitung. *Inovasi Matematika (Inomatika)*, 4(1). <https://doi.org/10.35438/inomatika.v4i1.302>
- Wahyuni, R., & Prihatiningtyas, N. C. (2020). Kemampuan Pemahaman Konsep Matematika terhadap Kemampuan Koneksi Matematika Siswa pada Materi Perbandingan. *Jurnal VARIABEL*, 3(2).
- Wati, N. M. (2018). *Analisis Kualitas Butir Soal Ujian Akhir Semester Gasal Standar Kompetensi Korespondensi Bahasa Indonesia Kelas X Kompetensi Keahlian Administrasi Perkantoran di SMK Negeri 1 Yogyakarta Tahun Ajaran 2017/2018* [Skripsi].
- Widoyoko, E. P. (2014). *Teknik Penyusunan Instrumen Penelitian*. Pustaka Belajar.
- Wulandari, S., & dkk. (2020). Pengembangan Soal Higher Order Thinking Skills (HOTS) pada Materi Aljabar di Sekolah Menengah Pertama. *Jurnal Didaktik Matematika*, 7(2). <https://doi.org/10.24815/jdm.v7i2.17774>
- Yusuf, A. M. (2015). *Asesmen dan Evaluasi Pendidikan Pilar Penyedia Informasi dan Kegiatan Pengendalian Mutu Pendidikan*. Prenadamedia Group.